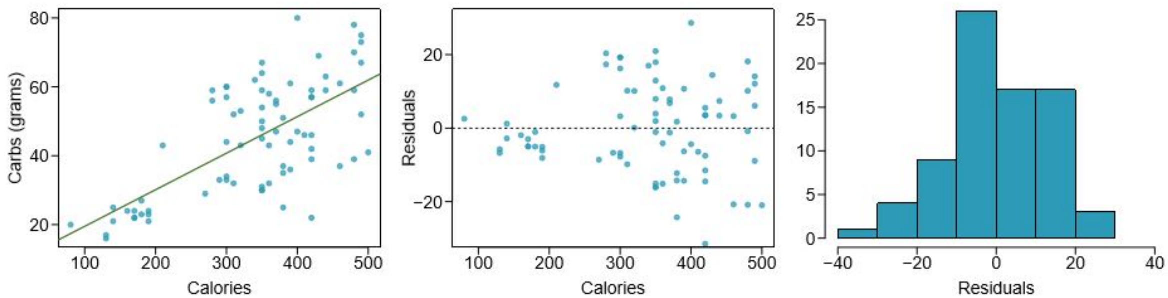
Chapter 7 HW

Christophe

November 3, 2015

7.24 Nutrition at Starbucks, Part I.

The scatterplot below shows the relationship between the number of calories and amount of carbohydrates (in grams) Starbucks food menu items contain. Since Starbucks only lists the number of calories on the display items, we are interested in predicting the amount of carbs a menu item has based on its calorie content



(a) Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.

The relationship between the number of calories and the amount of carbohydrates is a positive linear relationship. That is, as we see an increase in the number of calories, we also see an increase in the amount of carbohydrates.

(b) In this scenario, what are the explanatory and response variables?

Typically the x is the explanatory variable which is the calories and then typically the y is the response variable which is the amount of carbohydrates.

(c) Why might we want to fit a regression line to these data?

We would fit a regression line because the variables appear to have a positive linear relationship. This can be seen with the naked eye that the as the explanatory variable increase we see similar increases in the response variable.

(d) Do these data meet the conditions required for fitting a least squares line?

Yes, it appears to have linearity, the normal residuals appear to be nearly nomoral. There is constant variability and the observations are independent.

7.26 Body measurements, Part III.

Exercise 7.15 introduces data on shoulder girth and height of a group of individuals. The mean shoulder girth is 107.20 cm with a standard deviation of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

(a) Write the equation of the regression line for predicting height.

$\hat{height} = \beta_0 + \beta_1 * shoulder\ girth$

$b_1 = \frac{s_y}{s_x} R$

$b_1 = \frac{9.41}{10.37} * .67$

$b_1 = 0.61$

$y - \bar{y} = b_1(x - \bar{x})$

$y - 171.14 = -.61(x - 107.20)$

$y = -.61x - (.61 * 107.20) + 171.14$

$y = 105.748 + .61x$

$\hat{height} = 105.748 + 0.61_{shoulder\ girth}$

(b) Interpret the slope and the intercept in this context.

> The slope in this context is the amount of change in y over x. That is, for every 1 cm positive change in shoulder girth we would expect a .61 cm positive change in height. The intercept is somewhat meaningless in this context because we would not expect someone with a `0` shoulder girth to then have any height. It does provide some information regarding the limitations of the observations.

(c) Calculate R^2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

$R^2 = 0.67^2$

$R^2 = 44.89\%$

> In this context it means that `44.9%` of the variable of height can be explained by the variable of shoulder girth. That is there was a `44.9%` reduction in variablity in the line prediction of height by adding the variable of shoulder girth.

(d) A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model.

$\hat{height} = 105.748 + 0.61_{shoulder\ girth}$

$\hat{height} = 166.748$

(e) The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.

$e_i = y_i - \hat{y}_i$

$e_i = -6.748$

The residual of `-6.748` means we have a negative residual and it helps us to understand how well the line fits the data observation.

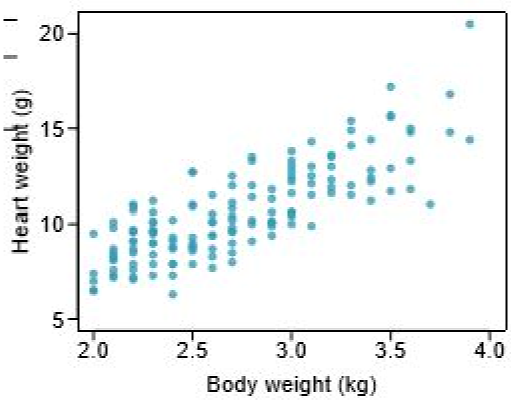
(f) A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child

I would not use this linear model to approximate the height. The reason being that our y-intercept is essentially 3.5 feet, since most one year olds will be under 3.5 feet using their should girth will only overstate their height and this model is not appropriate for this application.

7.30 Cats, Part I.

The following regression output is for predicting the heart weight (in g) of cats from their body weight (in kg). The coefficients are estimated using a dataset of 144 domestic cats.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.357	0.692	-0.515	0.607
body wt	4.034	0.250	16.119	0.000
$s = 1.452$	$R^2 = 64.66\%$		$R^2_{adj} = 64.41\%$	



(a) Write out the linear model.

$\widehat{cat_{heart\ weight\ grams}} = -0.357 + 4.034 * cat_{body\ weight\ kilograms}$

(b) Interpret the intercept.

The intercept of `-0.357` means that if a cat had 0 body weight then the heart weight would be negative. Since this is not realistic the intercept only serves to adjust the line to make it a better fit for the observations.

(c) Interpret the slope.

The slope of `4.035` indicates the change of y over x. So a 1 kilogram change in body weight would predict a 4.035 gram change in heart weight.

(d) Interpret R2.

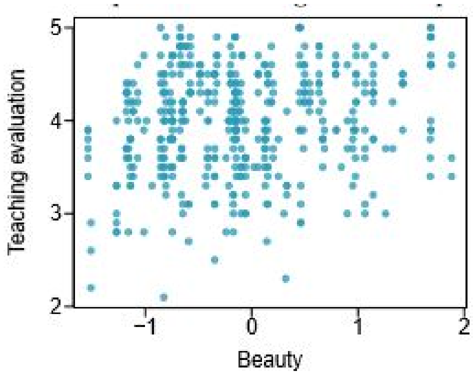
The R^2 of `64.66%` represents that the variation of height weight is `64.66%` explained by the change in cats body weight.

(e) Calculate the correlation coeficient

The correlation coefficient is 0.8041.

7.40 Rate my professor.

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. Researchers at University of Texas, Austin collected data on teaching evaluation score (higher score means better) and standardized beauty score (a score of 0 means average, negative score means below average, and a positive score means above average) for a sample of 463 professors. The scatterplot below shows the relationship between these variables, and also provided is a regression output for predicting teaching evaluation score from beauty score.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.010	0.0255	157.21	0.0000
beauty	$\boxed{}$	0.0322	4.13	0.0000

(a) Given that the average standardized beauty score is -0.0883 and average teaching evaluation score is 3.9983, calculate the slope. Alternatively, the slope may be computed using just the information provided in the model summary table.

$T = \frac{estimate - nullvalue}{SE}$

$\frac{x-0}{0.0322} = 4.13$

$\frac{x}{0.0322} = 4.13$

$x = 4.13 * .0322$

$b_1 = 0.132986$

(b) Do these data provide convincing evidence that the slope of the relationship between teaching evaluation and beauty is positive? Explain your reasoning.

Yes, the data does provide fairly convincing evidence that the slope of the teaching valuation is positive as the p value is `0.000` which is so low that we reject the null hypothesis that there is no correlation between teaching evaluation and beauty.

(c) List the conditions required for linear regression and check if each one is satisfied for this model based on the following diagnostic plots.

Linearity, the data appears to show a loose linear trend using the theoretical quantilies graph.

Nearly normal residuals, since there does not appear to be any significant outliers we can say the residuals are near normal using the residuals beauty chart.

Constant variability, the variability is roughly constant using the residuals bar chart.

Independent observations, based on the study described it would be acceptable to assume the observations are independent using the residuals order of data colletion chart.